The Casiquiare river acts as a corridor between the Amazonas and Orinoco river basins: biogeographic analysis of the genus *Cichla*

S. C. WILLIS,* M. NUNES,† C. G. MONTAÑA,‡ I. P. FARIAS,† G. ORTÍ§ and N. R. LOVEJOY¶

School of Biological Sciences, 348 Manter Hall, University of Nebraska-Lincoln, Lincoln, NE 68588, USA, †Laboratório de Evolução e Genética Animal, ICB, Universidade Federal do Amazonas, Estrada do Contorno 3000, Manaus, AM, Brazil, ‡Museo de Ciências Naturales, Universidad Nacional de Los Llanos, 3310 Mesa de Cavacas, Guanare, Venezuela, §Department of Biology, The George Washington University, 2023 G St. NW Suite 340, Washington, DC 20052, USA, ¶Department of Biological Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON, Canada M1C 1A4

Abstract

The Casiquiare River is a unique biogeographic corridor between the Orinoco and Amazonas basins. We investigated the importance of this connection for Neotropical fishes using peacock cichlids (*Cichla* spp.) as a model system. We tested whether the Casiquiare provides a conduit for gene flow between contemporary populations, and investigated the origin of biogeographic distributions that span the Casiquiare. Using sequences from the mitochondrial control region of three focal species (*C. temensis*, *C. monoculus*, and *C. orinocensis*) whose distributions include the Amazonas, Orinoco, and Casiquiare, we constructed maximum likelihood phylograms of haplotypes and analyzed the populations under an isolation-with-migration coalescent model. Our analyses suggest that populations of all three species have experienced some degree of gene flow via the Casiquiare. We also generated a mitochondrial genealogy of all *Cichla* species using >2000 bp and performed a dispersal-vicariance analysis (DIVA) to reconstruct the historical biogeography of the genus. This analysis, when combined with the intraspecific results, supports two instances of dispersal from the Amazonas to the Orinoco. Thus, our results support the idea that the Casiquiare connection is important across temporal scales, facilitating both gene flow and the dispersal and range expansion of species.

Keywords: *Cichla*, Neotropics, peacock bass, phylogeography, vicariance

Received 17 July 2008; revision received 13 December 2009; accepted 2 January 2010

Introduction

A fundamental task in evolutionary biology is determining the geographical context of speciation, and understanding how historical changes in geography have affected the divergence, dispersal, and extinction of lineages. The Casiquiare River of southern Venezuela is a unique biogeographic corridor between two of the most species-rich vertebrate faunas in the world (Reis *et al.* 2003; Winemiller *et al.* 2008). The Casiquiare provides a connection between the Amazonas, the world's

Correspondence: S. C. Willis, Fax: (402) 472 2083;
E-mail: swillis4@gmail.com

largest river system, and the Orinoco, the river system draining most of Venezuela and Colombia (Rice 1921) (Fig. 1). These two river systems contain extraordinarily diverse aquatic biotas and have long been considered separate biogeographic provinces and areas of endemism (Gery 1969; Weitzman & Weitzman 1982; Hubert & Renno 2006). Year-round, the Casiquiare captures flow from the headwaters of the Orinoco, but ultimately drains into the Negro River, the largest Amazonas tributary (Rice 1921; Lopez-Rojas *et al.* 1978). Thus, the Casiquiare appears to offer opportunities for both gene flow and dispersal between the aquatic biotas of the Amazonas and Orinoco (Mago-Leccia 1971; Freeman *et al.* 2007).
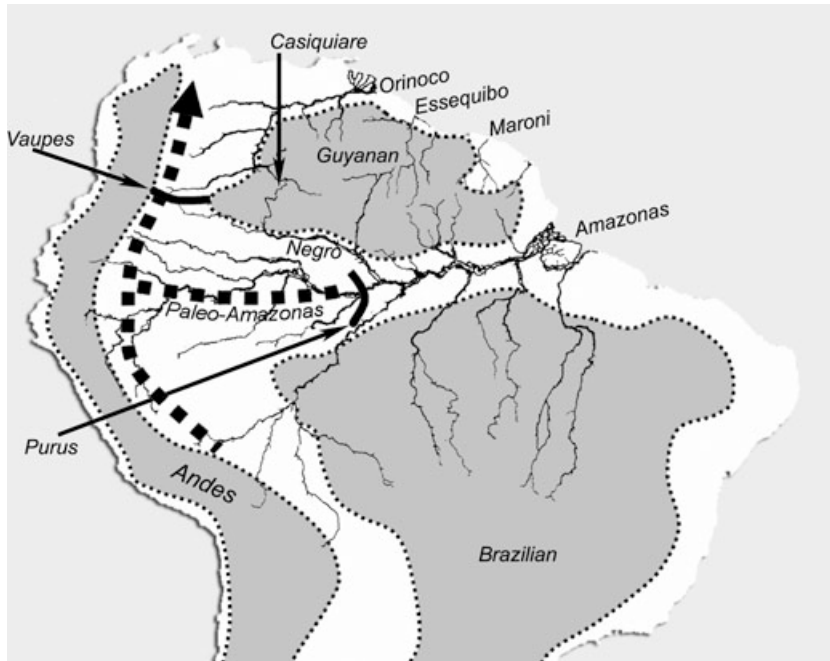
Fig. 1 Map of northern South America showing relevant geographic and geological features. Structural arches: Vaupes Arch; Purus Arch. Rivers: Orinoco, Essequibo, Maroni, Amazonas, Negro, Casiquiare. Major geological features: Guyanan Shield, Brazilian Shield, Andes Mountains. Tertiary River Drainage patterns: the 'Paleo-Amazonas' River (based on Lundberg *et al.* 1998b) that historically connected the Amazonas and Orinoco Basin regions.

The Casiquiare region has a complex and dynamic paleogeographic history (Lundberg *et al.* 1998b). Before ~10 Ma, a large 'Paleo-Amazonas' River likely flowed northwards along the eastern edge of the Andes, entering the Caribbean in the vicinity of contemporary Lake Maracaibo (Hoorn 1994). This ancient river drained areas now occupied by the upper (western) Amazonas and upper and western Orinoco, which presumably composed a single biogeographic region (Fig. 1). Indeed fossil fishes of many extant genera and species currently inhabiting the Amazonas, Orinoco, or both have been found in geological formations from the 'Paleo-Amazonas' period (Lundberg 1997, 1998a). Between 10–11 Ma, uplift in the Eastern Cordillera of the Andes caused the Vaupes Arch to come into closer contact with the Andes, forcing the 'Orinoco-Amazonas vicariance event' and separating the two drainages (Lundberg *et al.* 1998b). Subsequent foreland sedimentation from Andean erosion forced the Orinoco to shift east where it took up its current position along the western edge of the Guyanan shield (where the current Casiquiare connection lies), while the Amazonas eventually broke through its eastern barrier, the Purus Arch, to take up its current path to the mid-Atlantic (Bermerguy & Sena Costa 1991; Hoorn 1994; Hoorn *et al.* 1995). These movements provided subsequent opportunity for drainage capture between the Orinoco and Negro headwaters to the east of the Vaupes Arch, and at some point the Casiquiare connection was formed (Fig. 1). Unfortunately, the precise timing of the origin of the Casiquiare remains unknown (Stern 1970; Winemiller *et al.* 2008).

The Casiquiare region has likely played a role in both vicariance and dispersal of fishes between the Amazonas and Orinoco. The fragmentation of the 'Paleo-Amazonas' and the associated Amazonas-Orinoco vicariance event likely isolated previously widespread taxa, resulting in allopatric sister lineages. More recently, the origination of the Casiquiare River may have subsequently allowed dispersal (range expansion) or gene flow between the Amazonas and Orinoco (see Winemiller & Willis in pressfor a discussion of alternative regional corridors). Current fish distributions reflect both these possibilities (Winemiller *et al.* 2008). Some putative sister species, such as the piranhas *Pygocentrus natteri* and *P. cariba* have allopatric distributions in, respectively, the Amazonas and Orinoco; these distributions could be explained by the Amazonas-Orinoco vicariance event. Other species, such as the piranha *Serrasalmus rhombeus*, the catfish *Phractocephalus hemiliopterus*, and the Amazonas River dolphin *Inia geoffrensis*, have distributions that include the Orinoco, Casiquiare, and Amazonas (Winemiller *et al.* 2008). These taxa may owe their broad distributions to dispersal through the Casiquiare. For example, Freeman *et al.* (2007) hypothesized that dispersal from the Orinoco to the Amazonas (via the Casiquiare) explains the distribution of the widespread piranha species *Serrasalmus manueli*.

For species with ranges that currently encompass both the Amazonas and Orinoco, a natural expectation is that the Casiquiare River could allow gene flow between the two basins. However, limited phylogeographic data for Neotropical fishes suggest that this is

(a)

Cichla temensis

- ○ Orinoco
- ▣ Casiquiare
- ▲ Amazonas

(b)

0.005

(c)
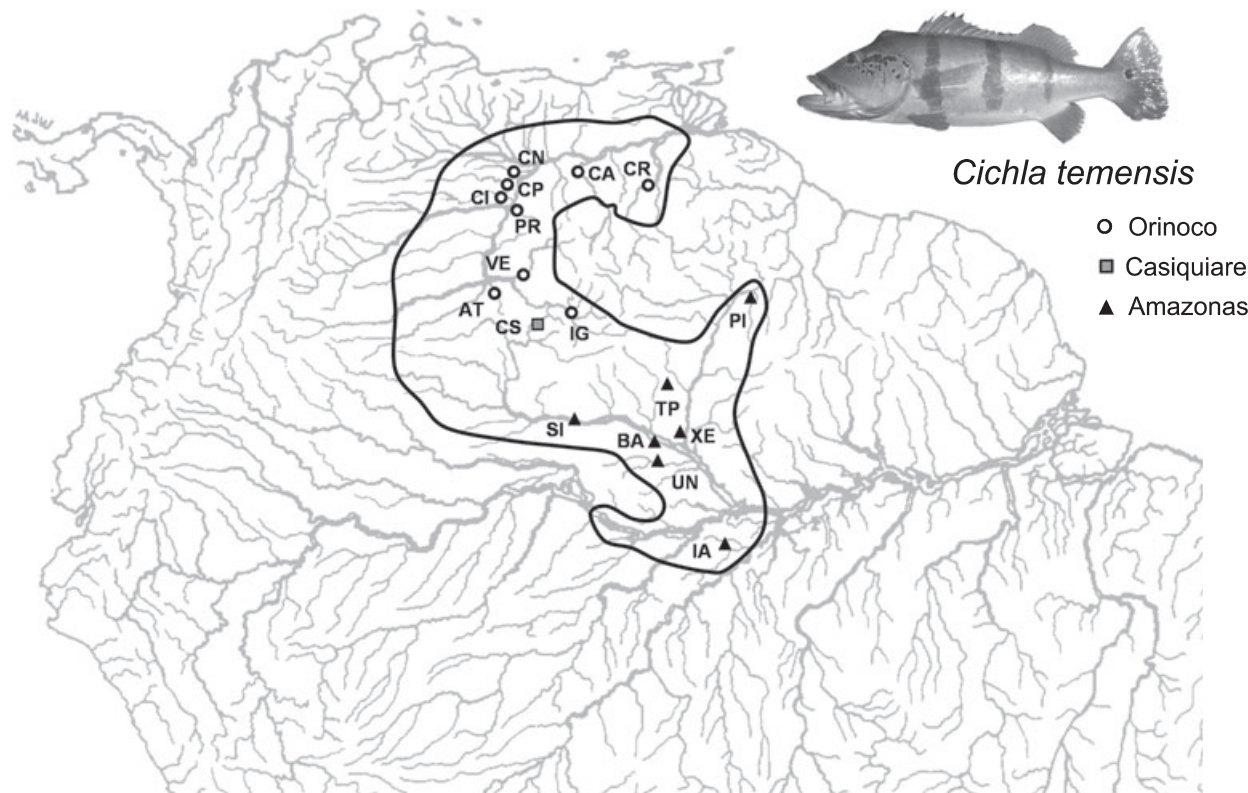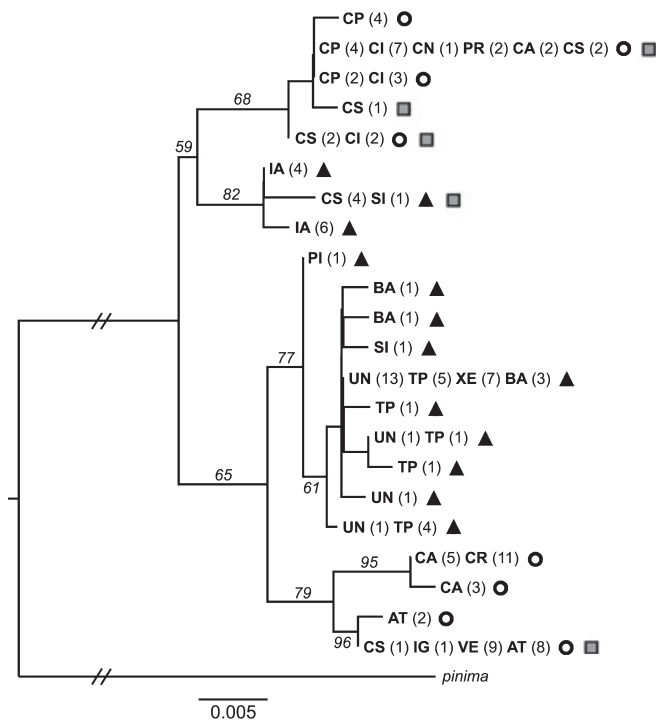
		model		
		full	**reduced**	**m=0**
Ori vs. Cas+Amaz	**Log(p)**	−9.2674 −8.2003	−11.6285 −10.6315	−8.6305 −8.6306
	t	1.62	1.75	0.026
	σ of *t*	0.001570	0.00012	0.000024
	θ_O	35400	39300	20400
	$m_{AC \to O}$	1.35×10^{-7}	3.85×10^{-6}	0
	θ_{AC}	51100	θ_O	76900
	$m_{O \to AC}$	6.44×10^{-6}	$m_{AC \to O}$	0
	θ_{OCA}	93	θ_O	86200
Ori+Cas vs. Amaz	**Log(p)**	−7.1003 −7.1254	−8.1246 −8.4007	−8.6490 −8.6488
	t	1.45	1.60	0.047
	σ of *t*	0.237	0.013312	0.000015
	θ_{OC}	45900	38100	32200
	$m_{A \to OC}$	3.64×10^{-6}	2.10×10^{-6}	0
	θ_A	51500	θ_{OC}	44200
	$m_{OC \to A}$	2.06×10^{-6}	$m_{A \to OC}$	0
	θ_{OCA}	28300	θ_{OC}	110500

not necessarily the case. Lovejoy & de Araújo (2000) inferred that the Casiquiare does not permit gene flow between populations of the needlefish *Potamorrhaphis guianensis* distributed in the Amazonas and middle and lower Orinoco. Similarly, Turner *et al.* (2004) found populations of *Prochilodus rubrotaeniatus* from the Orinoco and Amazonas are genetically distinct, and showed no evidence that mitochondrial haplotypes are shared between these two basins. Possible barriers reducing the importance of the Casiquiare as a corridor have been suggested. These include the Atures rapids of the upper Orinoco (Chernoff *et al.* 1991, Lovejoy & de Araújo 2000), and an environmental gradient from neutral, unstained clearwater in the upper Orinoco to acidic, tannin-stained blackwater in the upper Negro (Sioli 1984; Winemiller *et al.* 2008). Such barriers might also explain why some fishes, such as the arowana (*Osteoglossum* spp.), arapaima (*Arapaima gigas*), and discus cichlids (*Symphysodon* spp.) are present in the Amazon, but are not present in the Orinoco.

Here we investigate the biogeographic role of the Casiquiare, including its impact on both gene flow and species distributions, using the cichlid genus *Cichla*. *Cichla* is a clade of 15 species of large (upwards of 12 kg), colourful, and piscivorous fishes (Kullander & Ferreira 2006). Several species are economically important (Winemiller 2001), and are thought to play important roles in riverine food webs (Winemiller & Jepsen 1998; Layman & Winemiller 2004). *Cichla* is an ideal clade for examining the potential role of the Casiquiare because it includes species that are broadly distributed in all the drainages surrounding the Guyanan Shield. In addition, several species have distributions that span at least part of the Orinoco and Amazonas Basins and the Casiquiare. Previous work suggests that *Cichla* species diversification has been shaped by the complex historical evolution of South America's hydrography from at least the Miocene (Willis *et al.* 2007).

We used intraspecific genetic data from populations of three *Cichla* species found in the Amazonas and Orinoco in combination with a species-level phylogeny to evaluate the biogeographic role of the Casiquiare. We sought to answer the following questions: (1) Does the current Casiquiare canal facilitate gene flow between contemporary populations in each basin, or if not, how long ago did populations diverge? and (2) What is the historical biogeographic origin of the current distributions of Orinoco and Amazonas species? To answer these questions, we first use mitochondrial sequence data from multiple populations of the three species with distributions that span the Amazon, Orinoco, and Casiquiare (*Cichla monoculus*, *C. temensis*, and *C. orinocensis*) to assess patterns of genetic structure and haplotype sharing, as well as the timing of divergence between populations (for distributions of these species see Figs 2a, 3a, and 4a). Second, for a deeper temporal perspective on the origin of biogeographic patterns, we reconstruct a phylogeny for the genus *Cichla*, and use this topology for dispersal-vicariance analysis (DIVA; Ronquist 1997). This analysis should illuminate the role of the Casiquiare in the dispersal of evolutionary lineages. By combining phylogeographic and phylogenetic approaches, we provide a synthetic assessment of the biogeographic role of the Casiquiare.

Methods

Intraspecific analyses

For three focal species of *Cichla* (*C. temensis*, *C. orinocensis*, and *C. monoculus*), we obtained samples from localities throughout their distributions in the Amazonas, Orinoco, and Casiquiare (Figs 2a–4a; Table S1). Voucher specimens were regularly taken (voucher data available upon request), but due to logistical constraints, many specimens were photographed, sampled nondestructively (dorsal fin), and released alive.

For each of the three species, we collected 500–550 bp from the 5' portion of the mitochondrial control region (CR), using previously described primers and conditions (Willis *et al.* 2007). Sequences from 76 new samples (Genbank GU295709 – GU295740) were combined with sequences generated for our previous study (DQ841819-DQ841946). In this previous study (Willis *et al.* 2007) we identified several haplotypes in *C. monoculus* (from the Mavaca River in the Orinoco basin) and *C. orinocensis* (from three localities in the Negro Basin: Unini, Xeriuini, and Tapera) that we inferred to be introgressed from

Fig. 2 Intraspecific results for *Cichla temensis*. (a) Map of northern South America showing approximate species range and sampling sites for the current study. Locality abbreviations follow Table S1. (b) Maximum likelihood phylogram of haplotypes with bootstrap support proportions greater than 50%. For each haplotype, the locality, number of times observed, and drainage occupied are shown. (c) Results from IMa analysis. For each model and geographical grouping, parameters are as follows: log probability of the joint model (Log(p)), divergence time in millions of years (*t*), standard deviation of divergence time (σ of *t*) in millions of years, effective population size of Orinoco (θ_O) in individuals, migration rate of mtDNA from Orinoco to Amazonas ($m_{O \to A}$), effective population size of Amazonas (θ_A), migration rate of mtDNA from Amazonas to Orinoco ($m_{A \to O}$), and ancestral effective population size (θ_{OCA}). Models are: full model (six parameters), optimal reduced model, and model with migration forced to be zero. Values shown reflect a mutation rate of 6% sequence divergence per million years.
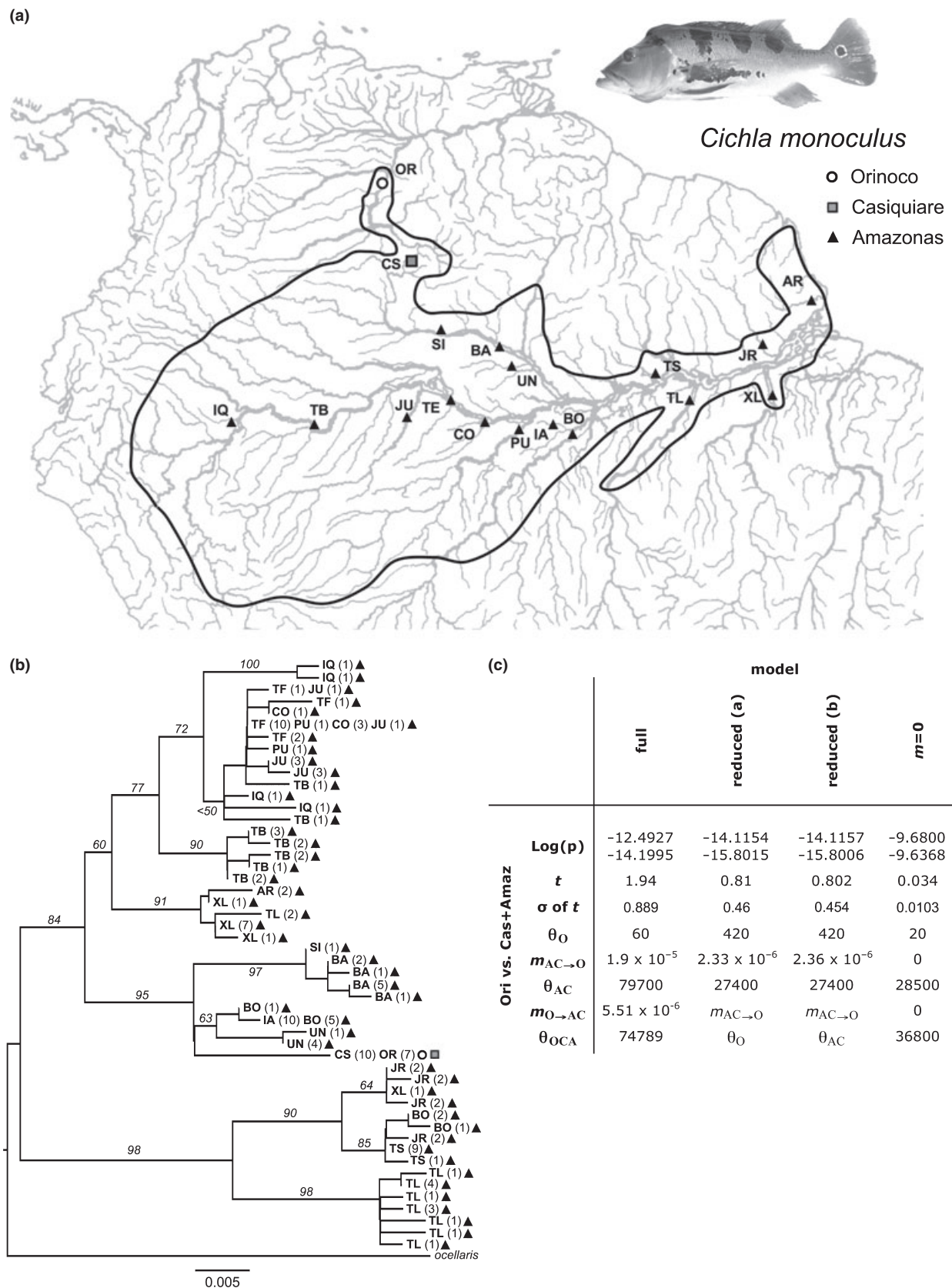
		model		
Ori vs. Cas+Amaz	**full**	**reduced (a)**	**reduced (b)**	**m=0**
Log(p)	−12.4927	−14.1154	−14.1157	−9.6800
	−14.1995	−15.8015	−15.8006	−9.6368
t	1.94	0.81	0.802	0.034
σ of t	0.889	0.46	0.454	0.0103
θ_O	60	420	420	20
$m_{AC \to O}$	1.9×10^{-5}	2.33×10^{-6}	2.36×10^{-6}	0
θ_{AC}	79700	27400	27400	28500
$m_{O \to AC}$	5.51×10^{-6}	$m_{AC \to O}$	$m_{AC \to O}$	0
θ_{OCA}	74789	θ_O	θ_{AC}	36800

Fig. 3 Intraspecific results for *Cichla monoculus*. As in Fig. 2.

(a)

Cichla orinocensis

- ○ Orinoco
- ▣ Casiquiare
- ▲ Amazonas

(b)

0.005

(c)

		model		
		full	**reduced**	**m=0**
Ori vs. Cas+Amaz	**Log(p)**	−7.3672 −7.5411	−10.0122 −10.2205	−9.6438 −9.6576
	t	1.91	1.78	0.067
	σ of t	0.0405	0.0276	0.00025
	θ_O	76100	67200	92800
	$m_{AC \to O}$	3.5×10^{-6}	2.38×10^{-6}	0
	θ_{AC}	55300	θ_O	60300
	$m_{O \to AC}$	1.55×10^{-9}	$m_{AC \to O}$	0
	θ_{OCA}	68100	θ_O	117600
Ori+Cas vs. Amaz	**Log(p)**	−9.0082 −9.2700	−10.6971 −10.9108	−10.2360 −10.2307
	t	1.87	1.78	0.066
	σ of t	0.0569	0.0217	0.00016
	θ_{OC}	74400	71900	91400
	$m_{A \to OC}$	3.33×10^{-6}	2.43×10^{-6}	0
	θ_A	62800	θ_{OC}	80200
	$m_{OC \to A}$	2.17×10^{-9}	$m_{A \to OC}$	0
	θ_{OCA}	39600	θ_{OC}	120900

Fig. 4 Intraspecific results for *Cichla orinocensis*. As in Fig. 2.

other species. These sequences were excluded from the following population level investigations.

For each of the three focal species, two separate alignments were generated using Clustal X (Thompson *et al.* 1997) with default parameters. The first alignment, used for phylogenetic (phylogeographic) analysis, used all unique CR sequences (haplotypes) for the focal species and one outgroup haplotype (selected based on Willis *et al.* 2007). The program Collapse (D. Posada-U. Vigo, Spain) was used to eliminate redundant sequences, with gaps treated as a fifth state and missing data ignored. The second alignment, used for population genetic (coalescent) analyses, included all sequences from the focal species, with no outgroup sequence. Data exploration suggested that modest variation in alignment parameters (gap opening/extension costs between 1 and 30) produced no change in recovered topologies, while higher alignment parameters produced biologically unrealistic alignments.

Phylogenetic analyses of haplotypes were conducted with Treefinder (Jobb *et al.* 2004) under maximum likelihood. In Treefinder, tree searches (1000 iterations) and bootstrap analysis (500 pseudoreplicates) were conducted under the model proposed by Treefinder. Treefinder was also used to execute SH tests of topologies constrained to fit different biogeographic scenarios (Shimodaira & Hasegawa 1999). We expected that reciprocally monophyletic groups of haplotypes in the Amazonas and Orinoco would indicate long-term isolation between populations in these two basins, while para- or polyphyletic haplotype lineages could suggest recent gene flow, or recent dispersal of a species from one basin to the other. Therefore, we tested for: (1) reciprocally monophyletic haplotype lineages in the Amazonas vs. the Orinoco, (2) monophyly of Orinoco haplotypes and unconstrained Amazonas haplotypes, and (3) monophyly of Amazonas haplotypes with unconstrained Orinoco haplotypes. Each of these three constraint tests was performed with haplotypes found exclusively in the Casiquiare treated as members of either the Amazonas or Orinoco groups, for a total of six comparisons to the unconstrained tree.

While reciprocal monophyly of haplotypes in the Orinoco and Amazonas basins would suggest long-term isolation between these regions, para- or polyphyletic groups could indicate either recent gene flow or incomplete lineage sorting between isolated populations. To discriminate between the latter two hypotheses, we used IMa, which implements a two-population isolation-with-migration model based on the coalescent (Hey & Nielsen 2007). The full IMa model has six parameters: population sizes for each of the two focal populations and the ancestral population, two migration parameters, and the time of divergence from a state of panmixia. In addition, using data from the full model, the program can examine nested models in which some of these parameters are equal (e.g. population sizes) or zero (e.g. migration) (we refer to these as 'reduced' models hereafter). These models, when not significantly less likely than the full model, are generally preferred, to avoid over-fitting the data to a complex model and increasing variance in resulting parameter estimates. The 'optimal reduced model' is the model with the fewest parameters that is not less likely than the full gene flow model. If reduced models with the migration parameters set to zero are not significantly less likely than the full model, we are unable to reject the null hypothesis of no gene flow between the Amazonas and Orinoco populations. For each species, data were analyzed using two models: (1) with migration (gene flow) permitted between populations and (2) with no migration ($m = 0$). Since IMa only allows for the analysis of two populations, we conducted separate analyses for *C. temensis* and *C. orinocensis* with Casiquiare individuals grouped with either the Amazonas or Orinoco populations (referred to below as geographical groupings). These combinations meant that four different sets of analyses were run for each of these two species. In the examinations with gene flow (the 'full' model, with six parameters), we also used the option in IMa to output the likelihoods of reduced models with fewer parameters. We compared the likelihoods of these to the full model using critical values from a chi-square distribution with appropriate degrees of freedom. In the results, we present the reduced models with the fewest number of parameters that were not significantly less likely than the full model. Unfortunately, it was not possible to examine reduced models in the no gene flow analyses, nor is it possible to directly compare these 'forced' zero migration models with the full or reduced models because of the Bayesian MCMC construction of IMa. For the broadly distributed *C. monoculus*, we used only Negro River haplotypes to represent the Amazonas population in IMa. This provided better consistency with assumptions of the IMa model (panmixia within populations), since samples from outside the Negro River represent distinct haplotype clades, indicating significant population substructure.

We analyzed our data under an HKY model in Bayesian runs constructed within the following ranges (depending on species): 100K to 200K genealogies sampled at an interval of 100 to 1000 generations with 20 to 30 chains under a geometric heating scheme, and following 3 to 10 million burn-in generations. For each analysis, numerous preliminary runs were made to determine the correct prior maxima and Bayesian conditions (heating scheme, length of burn-in, etc.) before two separate final runs were performed using different

number seeds. From these final runs, we recorded the parameter estimates for the joint model. Analyses were constructed to ensure that effective sample sizes in the final runs were above 50, and where possible, posterior distributions were fully encompassed by prior maxima (see Results). To convert parameter values to meaningful biological estimates, we used a mutation rate for the CR calibrated in African cichlids [6%/Myr; derived from the divergence of two Lake Malawi cichlid clades (Sturmbauer *et al.* 2001) and confirmed in Lake Victoria cichlids (Verheyen *et al.* 2003)] and a generation time of 5 years (Winemiller pers. comm.).

Cichla phylogeny and dispersal vicariance analysis

We evaluated *Cichla* phylogeny using three mitochondrial loci: CR, cytochome *b* (cyt *b*) and ATPase 8 and 6 (ATP). Portions of CR (500–550 bp) and cyt *b* (670 bp) were amplified and sequenced using primers and parameters described in Willis *et al.* (2007). Building on the dataset from Willis *et al.* (2007), we collected cyt *b* and CR sequence data for 102 new individuals, representing all remaining described species of *Cichla* (*C. mirianae*, *C. thyrorus*, *C.* cf. *vazzoleri*, *C. jariina*, *C. kelberi*, and *C. piquiti*), as well as additional localities for several species. New sequences have been deposited in Genbank as GU295666-GU295690 and GU295691-GU295708.

We also collected a new ATPase 8,6 dataset for *Cichla*. PCR primers for ATP were designed for this study: ATPLabF (5′ AGCGTTAGCCTTTTAAGC 3′) and ATPLabR (5′ ACTATGTGGTATGCGTGTGC 3′). For amplification, 25 µL reaction volumes contained 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl$_2$, 150 µM each dNTP, 0.4 µM each primer, 0.5 U of Invitrogen recombinant *Taq* polymerase, and 1 to 4 µL DNA extract (∼10–50 ng/µL). Touchdown PCR was performed with an initial denaturation step at 94 °C for 30 s, followed by 30 cycles of 45 s at 94 °C, 60 s at *X* °C, and 90 s at 72 °C, then a final extension of 5 min at 72 °C, where *X* was 58 °C for 5 cycles, 56 °C for 5 cycles, 54 °C for 5 cycles, and 52 °C for 20 cycles on an MJ Research PTC200 thermal cycler. PCR products were sequenced at the University of Washington High Throughput Facility. New sequences have been deposited in Genbank as GU295741-GU295801.

The resulting dataset includes >2000 bp for 148 individuals representing all of the species recognized in Kullander & Ferreira's (2006) review of *Cichla,* although here we consider *C. nigromaculata* synonymous with *C. monoculus* (see Fig. S1 and Table S2 for sampling sites and species distributions). The cyt *b* and ATP sequences were trivial to align as no insertion-deletion events were detected. Individuals of each species with redundant sequences at cyt *b* and ATP were noted and eliminated from the matrix, leaving 73 unique haplotypes, and these were concatenated with their CR sequence. The CR haplotypes were aligned as described above. Concatenation of sequences from these regions is appropriate since mtDNA is inherited as a single, nonrecombining unit and all sub-regions experience the same evolutionary history regardless of mutation pattern (Meyer 1993). The final matrix was submitted to TreeBase (SN4802).

To infer a genealogy of *Cichla* mitochondrial haplotypes, tree searches were implemented under parsimony (MP) and maximum likelihood (ML) criteria in PAUP* 4b10 (Swofford 2000) and Bayesian likelihood analysis (BA) in MrBayes v. 3.1.2 (Ronquist & Huelsenbeck 2003). Models for each partition were chosen using ModelTest 3.7 (Posada & Crandall 1998) or MrModeltest (Nylander *et al.* 2004) for ML and BA respectively. Heuristic searches for ML included 1000 random addition replicates with unlimited tree bisection and reconnection branch swapping (TBR). This analysis was run with a single partition that included all data. Because MrBayes implements data partitioning more easily than PAUP*, BA was run with separate models of sequence evolution applied to each of 10 partitions corresponding to each gene (4, as ATP8 and 6 have different reading frames), and one each for 1st, 2nd, and 3rd codon positions within the three protein-coding genes. For the Bayesian analysis, two simultaneous runs were executed for 20 million generations with eight chains under default heating, sampling every 1000th generation, and the first 19 000 samples were discarded as burn-in. Convergence of runs in MrBayes was estimated using Tracer (http://evolve.zoo.ox.ac.uk/). MP analyses included a heuristic search with 10 000 random addition search replicates with TBR. Support for branches was estimated using 1000 bootstrap replicates (Felsenstein 1985), each with 100 random-addition heuristic searches limited to 10 000 rearrangements per random-addition replicate.

We inferred a species-level phylogeny by implementing the Minimize Deep Coalescences (MDC) method (Maddison & Knowles 2006), using the Deep Coalescences module in Mesquite 2.6 (Maddison & Maddison 2008). The Deep Coalescences module estimates the minimum number of gene lineages that must be postulated as deep coalescences (resulting from incomplete lineage sorting or minor ongoing gene flow) for a gene tree contained in a particular species-level topology, and allows Mesquite to commit heuristic searches of species trees that minimize the number of deep coalescences. As gene tree input, we used the maximum credibility tree from the BA (i.e. the tree with the greatest product of posterior probabilities), and instructed Mesquite to ignore branch lengths.

To generate probable historical biogeographic scenarios for *Cichla*, we used a dispersal-vicariance analysis (DIVA; Ronquist 1997). DIVA estimates the possible character states (distribution) of each ancestral node in the phylogeny by parsimony optimization, where divergence events between areas (vicariant speciation) and within areas (sympatric or microallopatric speciation) have a cost of zero, while dispersal or extinction events have a cost of one. Unlike other techniques to estimate ancestral distributions, DIVA does not force the distributions of ancestral species among geographical areas to be mutually exclusive, meaning that ancestral species can be distributed in more than one area at a time, as can contemporary species. As input, DIVA requires a fully-resolved species phylogeny and a distribution matrix of presence or absence of tip taxa in each geographic area. For the present analysis, species were scored as present or absent in areas defined by river basin: the Amazonas (including all tributary rivers), Orinoco, or Guyanas (including both the Essequibo and Maroni Rivers) (see Fig. S1).

Results

Intraspecific analyses

Cichla temensis. We obtained 546 aligned bp of CR from 128 individuals of *C. temensis* from seven localities in the Amazonas basin, nine localities in the Orinoco basin, and one locality in the Casiquiare (Fig. 2a). These sequences collapsed to 22 haplotypes, none of which were shared between the Orinoco and Amazonas: 12 haplotypes were distributed exclusively in the Amazonas, and five were distributed exclusively in the Orinoco. Five haplotypes were found in the Casiquiare, of which one was exclusive to this location, three were shared with the Orinoco, and one was shared with the Amazonas. Treefinder proposed a TN+Γ model of evolution, and phylogenetic analysis resulted in a tree of likelihood −1139.329 (Fig. 2b). In this phylogeny, haplotypes distributed in the Amazonas and Orinoco basins did not form reciprocally monophyletic groups, although haplotypes from the same or nearby localities were often clustered together. The haplotype unique to the Casiquiare was most closely related to a clade of haplotypes found in both the Orinoco and Casiquiare. When this exclusive haplotype was grouped with Orinoco haplotypes, SH topology tests showed that none of the three tested topologies (reciprocally monophyletic drainages, monophyletic Amazonas, monophyletic Orinoco) were significantly less likely than the geographically reticulate one we recovered by unconstrained heuristic search (*P* = 0.1966, 0.2039, 0.1745, respectively). However, each of these constraints was less likely than the unconstrained topology when the Casiquiare-exclusive haplotype was constrained to group with Amazonas haplotypes (*P* = 0.0235, 0.0227, 0.0216). Overall, the phylogeographic evidence suggests a history of genetic exchange between the Orinoco and Amazonas, presumably via the Casiquiare, as this locality exhibited haplotypes distributed across the tree and shared with both two major basins. However, no haplotypes were shared between basins outside of the Casiquaire River.

Our IMa analysis showed that reduced models that did not include gene flow were significantly less likely than the full model that did (*P* < 0.05) (Fig. 2c). IMa indicated that a reduced model (with equal population sizes and equal and non-zero migration rates) was not significantly less likely than the full model for both geographical groupings [2× log-likelihood ratio (2LLR) = 2.0486, 2.5505, 4.8532, 4.8623 with 3 d.f.]. The alternative geographical grouping of individuals from the Casiquiare locality with either the Amazonas or Orinoco populations did not significantly alter the comparative likelihood of gene flow models. These results suggest that genetic exchange between the Amazonas and Orinoco is a better explanation for observed haplotype sharing than ancestral polymorphism. However, we note that we were unable to establish prior maxima that fully encompassed the posterior distribution of the time parameter in the full model for either geographical grouping; rather, the posterior distributions rose sharply to a plateau at approximately 400–500 kya but then remained asymptotic over higher values. This is evident in the standard deviation associated with these values (Fig. 2c) and suggests convergence to an equilibrium island model with open-ended divergence time (Table S3); however, the minimum divergence time that is consistent with the estimated rate of gene flow between these populations is 400–500 kya (see Discussion). Joint model analysis in IMa suggested that the divergence time between Amazonas and Orinoco populations was 1.45 to 1.75 Ma (with ongoing migration thereafter). Alternatively, without gene flow, divergence times were much younger (26 kya–47 kya), indicating that if we exclude gene flow from the model used to estimate gene genealogies, it is necessary to infer that populations were connected very recently. All parameters were fully bounded in the no-migration analyses.

Cichla monoculus. For *Cichla monoculus*, we obtained approximately 520 bp from 139 individuals from 16 localities in the Amazonas basin, one locality in the Orinoco basin, and one locality in the Casiquiare, reflecting the proportional distribution of this species in both basins (Fig. 3a). The sequences collapsed to 50 haplotypes. Most of these haplotypes were found in the Amazonas basin; in fact, only a single *C. monoculus* haplotype was found in

the Orinoco basin, and it was shared with all fishes from the Casiquiare locality. Treefinder suggested a TN+Γ model, and phylogenetic analysis produced a tree of likelihood −1515.904 (Fig. 3b). In this tree, the haplotype from the Orinoco and Casiquiare was nested deeply within the Amazonas haplotypes, suggesting a recent divergence or colonization. A topology where this haplotype was constrained to be sister to a clade containing the Amazonas haplotypes was significantly less likely than the recovered tree (SH test, $P = 0.0285$). The asymmetric distribution of genetic diversity between basins, and the topology of the haplotypes, suggests a recent colonization of the Orinoco basin by *C. monoculus* from the Amazonas basin.

For IMa, we excluded analyses with the geographical grouping Orinoco+Casiquiare because this appeared to violate the IMa assumption of no reciprocal monophyly between populations. Analysis of the Casiquiare + Negro geographical grouping with IMa showed that a reduced model with no migration was significantly less likely than the full model, falsifying the hypothesis of no gene flow (Fig. 3c). We found that two reduced models were not significantly less likely than the full model (2LLR = 3.2452 and 3.2040 for model (a), and 3.2459 and 3.2022 for model (b), both with 2 d.f.), but both of these models contained gene flow between the Orinoco and Amazonas populations. In fact, all models without bi-directional gene flow were less likely than the full model ($P < 0.05$). We were able to establish a maximum prior for divergence time in this analysis; this parameter had a unimodal distribution with a peak between 40 kya and 75 kya in both runs, but also had a wide right tail that increased the average. However we experienced difficulty finding an appropriate maximum prior for Amazonas (Negro) population size. The marginal distribution of this parameter showed a distinct unimodal peak that was fully encompassed, but the upper tail of the distribution reached a plateau at low but non-zero probability values for higher values of population size. Further, despite the peak in the marginal distribution, the joint estimate of divergence time between Amazonas (Negro) and Orinoco populations was 1.94 Ma and ∼800 kya in the full and reduced models respectively. In contrast, we found that when migration was set to 0, divergence time estimates were relatively recent (∼34 kya), albeit with a wide variance (Fig. 3c). Very small effective population sizes were estimated for the Orinoco population which is not surprising given that a single control region haplotype was shared by all 17 fishes.

Cichla orinocensis. We sequenced 550 bp from a total of 105 individuals of *Cichla orinocensis* from two localities in the Amazonas basin, nine localities in the Orinoco

basin, and one locality in the Casiquiare (Fig. 4a). The sequences collapsed to 26 haplotypes. Five of these were found only in the Amazonas basin, while 19 were only found in the Orinoco. The 10 Casiquiare fishes sequenced had two haplotypes: one of which was shared with the Amazonas (Negro), while the other was shared with the Orinoco. No shared haplotypes were observed between the Orinoco and Amazonas. Implementing a HKY+Γ model of evolution returned a tree of −1221.071 (Fig. 4b). In this tree, Amazonas and Casiquiare haplotypes were nested among Orinoco haplotypes, but this topology was not significantly more likely than any of the constrained topologies ($P > 0.2$).

In IMa analysis of *C. orinocensis* data, models that did not include migration were significantly less likely than the full model ($P < 0.05$). The optimal reduced model had equal population sizes and migration rates, in both runs of both geographical groupings (2LLR = 3.3777, 3.2816, 5.2900, 5.3588 with 3 d.f.; Fig. 4c). However, as in *C. temensis*, we were unable to find an effective maximum prior for divergence time in the full model. This parameter rose asymptotically to a peak between 650 kya and 750 kya for both geographical groupings and then stayed higher at increasing values of *t*. Consequently, divergence time estimates from the joint models (considering all parameters simultaneously across genealogies) with gene flow ranged from 1.78 to 1.91 Ma (with a wide variance), where as divergence times without gene flow were much younger (∼67 kya) with low variance.

Species phylogeny

After removal of redundant OTUs, our matrix for species phylogeny included 73 *Cichla* mitochondrial haplotype OTUs, each 2035 bp (670 cyt *b*, 158 ATP 8, 683 ATP 6, 524 CR). Maximum uncorrected sequence divergence among the *Cichla* haplotypes was approximately 7% for cyt *b*, 9% for ATP, and 14% for CR. In the alignment, 613 bp were variable and 434 were parsimony informative. Based on the relative rates estimated in the BA, CR had the highest estimated rate of molecular evolution, followed by 3rd, 1st, and 2nd codon positions across protein-coding genes.

Phylogeny estimation using MP, ML, and BA all recovered very similar trees (Fig. 5), each with the same general topology as our previous analysis (Willis *et al.* 2007). ML recovered a tree with likelihood −9369.31448 under the GTR+I+Γ model chosen by Modeltest. BA found a model and tree with marginal likelihood −8757.94 using the following models for the 10 partitions chosen by MrModeltest (K80, JC, HKY, GTR, HKY, GTR+I, K80 + I, F81, HKY+Γ, HKY+I+Γ). MP

Fig. 5 Maximum credibility phylogram of *Cichla* haplotypes based on 2035 bp of the mitochondrial control region, ATPase 8,6 and cytochrome *b* loci. Values associated with branches are parsimony bootstrap and Bayesian posterior probabilities respectively. Branches marked with an asterisk had 99 or 100% for bootstrap values, and 1.0 for posterior probability.

supported 2016 equally-parsimonious trees, each with 1289 steps. These MP trees varied exclusively in the topology of haplotypes within species, with one exception. All three analyses indicate that *Cichla* species are partitioned into two main clades. The first of these (clade A) was comprised of *Cichla temensis*, *C.* cf. *vazzoleri*, *C. pinima*, *C. thyrorus*, *C. mirianae*, *C. melaniae*, *C. piquiti*, and *C. jariina*. Within clade A, some species did not exhibit monophyletic haplotype clades, although in only one case was a haplotype shared among species at all three loci (*C. pinima* and *C. jariina*). While the most common topology in the BA (PP 0.38) placed *C. jariina*, *C. piquiti*, and part of *C. pinima* as basal in Clade A, no analysis provided strong support for this arrangement: MP trees disagreed, and the optimal ML topology was a trichotomy. However, in no analysis was *C. temensis*

at the base of the clade. The second clade of *Cichla* (clade B) contained two well-supported subclades: clade B1, including *C. monoculus*, *C. ocellaris*, *C. kelberi*, and *C. pleiozona*; and clade B2, including *C. orinocensis* and *C. intermedia*. The monophyly of clade B was well supported by MP bootstrap analysis, but only modestly supported by posterior probability. In addition, different methods produced minor rearrangements of closely related haplotypes, but these produced no effect on the subsequent biogeographic analyses.

Based on our mtDNA genealogy, a few species were not characterized by monophyletic clades of haplotypes (*C. pinima*, *C. mirianae*, *C. melaniae*, and *C. ocellaris*) that we interpret to result from incomplete lineage sorting (see Willis *et al.* 2007 for a discussion). To convert the genealogy to a fully-resolved species-level phylogeny

for use in DIVA, we used the Minimize Deep Coalescences module of Mesquite. MDC Analysis of the maximum clade credibility Bayesian tree produced 20 trees with 17 deep coalescences. We suspected that uncertainty regarding the placement of *C. pinima* was resulting in multiple topologies, so we removed this taxon and repeated the analysis. This resulted in a single tree, to which we added *C. pinima* as indicated by the majority of its haplotypes (Fig. 6a). As discussed below, our biogeographic conclusions are unaffected by this selection of the final tree or any of the 20 initially produced.

Dispersal-vicariance optimization

Dispersal-vicariance analysis of the species phylogeny resulted in four equally parsimonious optimizations, summarized in Fig. 6. Each optimization requires five dispersal events and/or extinctions (these are equivalent in DIVA, much like accelerated and delayed transformation in parsimony optimizations). In each optimization, the inferred geographic distribution for clade A is identical: all internal nodes are reconstructed as Amazonas. For this clade, a single dispersal event into the Orinoco was inferred, explaining the current distribution of *C. temensis* (Amazonas and Orinoco). As mentioned above, this reconstruction is unaffected by alternative phylogenetic arrangements recovered in different species trees, as long as *C. temensis* is not placed as the basal species in clade A. In none of the 20 initial MDC species trees was this observed.

In clade B, different possible histories were recovered for *C. monoculus*. Two optimizations (Fig. 6a and c) indicate that this species dispersed from the Amazonas to the Orinoco after a vicariant divergence of its ancestor between the Amazonas and Guyanas. The scenario implied by the remaining two optimizations (Fig. 6b and d) is that the common ancestor of *C. monoculus* and *C. ocellaris* was distributed in the all three biogeographic provinces (Amazonas, Orinoco, and Guyanas), and that vicariance separated *C. monoculus* (Amazonas, Orinoco) from *C. ocellaris* (Guyanas).

In the case of *C. orinocensis*, two optimizations (Fig. 6a and b) indicate that *C. orinocensis* dispersed from the Orinoco to the Amazonas after a 'non-vicariant' divergence of the common ancestor of *C. orinocensis* and *C. intermedia* in the Orinoco. Alternatively, *C. orinocensis* inherited its broad distribution from an ancestor that dispersed from the Amazonas to Orinoco (Fig. 6c), or *C. orinocensis* dispersed directly from the Amazonas to the Orinoco (Fig. 6d).

Interestingly, two different optimizations were determined for the common ancestor of clade B. Two optimizations (Fig. 6a and b) suggested a history in which the ancestor of clade B species dispersed from the Amazonas to the Orinoco, and the origin of clades B1 (*C. monoculus et al.*) and B2 (*C. orinocensis* & *C. intermedia*) resulted from vicariance between the Amazonas and Orinoco. The other two optimizations (Fig. 6c and d) implied a scenario in which the common ancestor of clade B was distributed only in the Amazonas, with
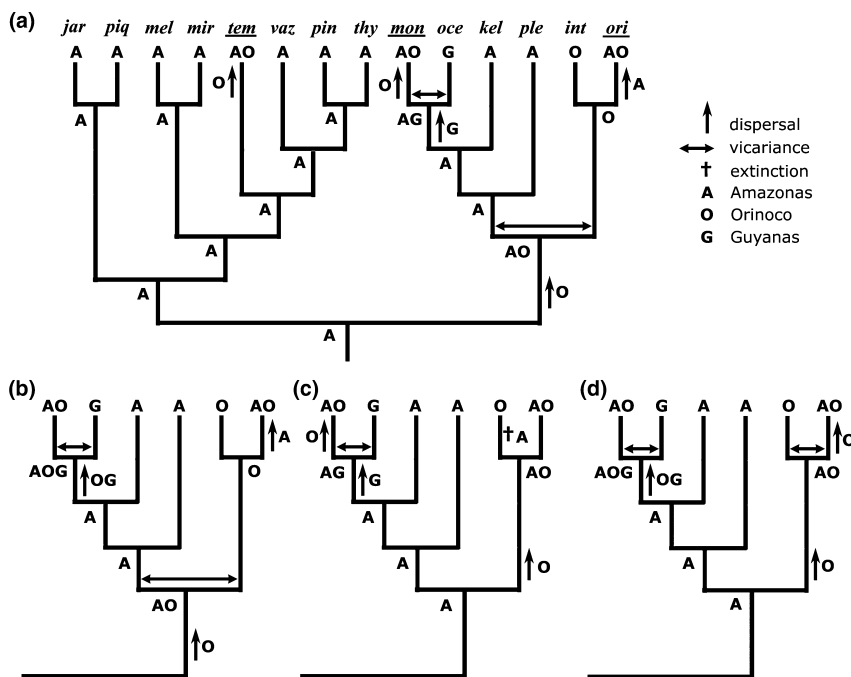

Fig. 6 Results of the Dispersal-Vicariance Analysis based on the *Cichla* species phylogeny inferred using Minimize Deep Coalescences. For each extant species, distribution in the Amazonas basin, Orinoco basin, and/or Guyanas drainages (e.g. Essequibo) is indicated, and focal species are underlined. Inferred distributions for ancestral species are shown at internal nodes. Dispersal, vicariance, and extinction events are shown on respective branches, and the order of events proceeds from bottom to top along each branch. Only the right portion of the phylogeny, Clade B, is shown for scenarios b through d. The optimization for Clade A is identical in all cases.

lineages from each sub-clade dispersing into the Orinoco independently.

Discussion

The Casiquiare River is a uniquely important biogeographic phenomenon because it provides a potential connection between two of the world's most species-rich vertebrate faunas (Reis *et al.* 2003). The origin of these faunas resulted not only from vicariance and isolation of lineages in the Amazonas and Orinoco basins, but also from exchange and interactions of lineages in secondary contact. Phylogeographic examination of genetic diversity provides a powerful tool for reconstructing species histories and the impacts of historical alterations of river drainage connectivity (e.g. Burridge *et al.* 2006, 2007). Here, mtDNA sequences from three species of cichlid fishes (*Cichla* spp.) provide strong evidence that the Casiquiare River has acted as a corridor for gene flow and dispersal between the Amazonas and Orinoco basins.

For the three species of *Cichla* examined, we discovered haplotype sharing across drainages, as well as poly- or paraphyletic distributions of haplotypes across drainages, implying relatively recent co-ancestry. For instance, in *C. monoculus* all fishes in the Orinoco basin possessed a single haplotype that was shared by all fishes in the Casiquiare. Further, this haplotype was closely related to haplotypes from the Negro and lower Madeira Rivers, and nested deeply within the remaining genetic structure of this species. This pattern suggests a recent dispersal of *C. monoculus* from the Negro into the Orinoco basin. In contrast, *C. temensis* exhibited complex genetic structure across basins, with some haplotypes more closely related to those in the opposing basin rather than those at nearby localities. This suggests a deeper history of population divergence with multiple instances of migration between basins. However, the relative importance of divergence and migration is difficult to ascertain from phylogenies alone.

To distinguish between ongoing gene flow and incomplete lineage sorting as explanations for the distribution of genetic diversity, we applied coalescent analyses as implemented in the IMa program. We found that optimal reduced models for each species included gene flow between populations in the Amazonas and Orinoco drainages. Although we had difficultly in estimating divergence time in several cases because the analyses began converging on equilibrium-island models, we found that when we did not consider gene flow divergence time was much younger. Joint parameter estimates ranged from 800 kya to 1.91 Ma in the full and reduced models. In contrast, estimates from models without gene flow were less than 70 kya in all cases

(Figs 2c, 3c, and 4c). This difference in estimated divergence times between gene flow and no gene flow models follows from coalescent theory: if no gene flow is allowed between populations, any sharing of recently derived haplotypes must be explained by inheritance from the ancestral population (incomplete lineage sorting), a pattern promoted by recent divergence times and/or large effective population sizes. However, when gene flow is allowed, sharing of gene lineages can be analytically divided between inheritance and post-divergence genetic exchange, and divergence times become correspondingly older. In many cases though, and particularly where only one or a few loci are analyzed, it is possible for estimated gene flow to be so high that this masks the signature of inheritance from an ancestral population, and divergence times become effectively infinite (an equilibrium island model). In these cases, where divergence time is asymptotic, the mode of divergence time (unimodal distribution) or the approximate value at which the asymptote is first reached (when high divergence times have equal probability) can be informative. These values represent the minimum divergence times that are consistent with the estimated rates of gene flow. In the present study, these values were much lower than the average or joint model estimates (420–460 kya for *C. temensis*, 680–730 kya for *C. orinocensis*, and 40–75 kya for *C. monoculus*), and much more similar to estimates made with the no gene flow models.

Our analysis of population structure using Isolation-with-Migration (IMa), has some shortcomings. The first is our use of a single genetic marker (mtDNA). It has been established that coalescent analyses are more accurate when based on multiple independent loci, and any single locus could provide a biased portrait of gene flow and population structuring, such as in the case of sex-biased migration (e.g. Cooper *et al.* 2009). Ongoing studies with nuclear markers should reveal whether this is the case. Second, our analysis may have violated some of the assumptions of the isolation-with-migration model, principally the assumption of panmixia (no geographic structure) in each population. Population structure has the effect of preserving genetic diversity despite a reduced overall effective population size (sensu Wright 1943). In a coalescent analysis, this pooled genetic diversity may have the effect of inflating estimated effective population sizes (i.e. the Wahlund effect, Wahlund 1928; see also Harrison & Hastings 1996; Lacy 1987). It is unclear exactly what effect such inflated population sizes would have in an isolation-with-migration analysis, but recent studies have found IMa to be relatively robust to violation of the assumption of panmixia (Strasburg & Rieseberg 2009). These inflated population sizes may act to increase the

probability of incomplete lineage sorting and decrease divergence time inferred under the current dataset, although the relative size of these effects may depend on the particular analysis and mutational distribution. Third, we were able to include relatively few localities in some portions of the range of each species. The Casiquiare, upper Orinoco, and upper Negro are remote and difficult to sample regions, limiting our ability to collect additional samples. However, our study is the first to examine species distributed across this region in a direct investigation of gene flow using parametric methods.

The shortcomings described above may have limited our ability to fully bound certain parameters with prior maxima or resulted in wide posterior distributions. Nevertheless, we believe that this does not invalidate important conclusions from our analysis. In the likeliest case, true rates of gene flow and divergence times lie between those estimated with each model variation (e.g. Peters *et al.* 2005; Hansson *et al.* 2008; Mäkinen & Merilä 2008). Thus, although we may not be able to rely on precise estimates of divergence time or quantitative rates of gene flow in either model, these results confirm that the Casiquiare region either serves as a conduit for contemporary gene flow between the Amazonas and Orinoco, or that gene flow across this region ceased very recently. Interestingly, the estimated rates of migration between basins are relatively low, on the order of a few alleles effectively exchanged per hundreds of thousands or million years (Figs 2c, 3c, and 4c). Based on the distribution of haplotypes in the phylogeny and in space, it is clear that this level of gene flow has not prevented regions in each basin from developing their own endemic set of mitochondrial haplotypes (Figs 2b, 3b, and 4b). However, in most cases these haplotype groups are not reciprocally monophyletic by basin, and are separated by only a few mutations at a rapidly-evolving locus, suggesting that gene flow was an important factor in the origin of this distribution.

Our intraspecific results indicate that the Casiquiare River has likely been responsible for gene flow between the Amazonas and Orinoco basins. This contrasts with previous studies that found no evidence for recent genetic transfer across this region (Lovejoy & de Araújo 2000; Turner *et al.* 2004). However, intraspecific analyses provide limited information regarding the formation of species distributions and biogeographic patterns over longer timescales. The question of whether lineages have dispersed between basins by using the Casiquiare and in what direction, as well as the possible effects of the Amazonas/Orinoco vicariance event, are best answered using our DIVA results.

Our recovered species phylogeny agrees well with our previous results (Willis *et al.* 2007), but includes species and localities that were not available for that analysis. Different analytical methods recovered very similar phylogenetic trees, indicating that the signal for most nodes is quite strong. The dispersal-vicariance analysis (Fig. 6) provided unambiguous evidence that *C. temensis* dispersed from the Amazonas basin into the Orinoco. This finding is based on the exclusively Amazonas distribution of the clade A relatives of *C. temensis*. This scenario is reasonably consistent with the intraspecific data for this species (Fig. 2), which show a polyphyletic distribution of mtDNA lineages between basins. However, the complex geographic structure of relationships between haplotypes across drainages suggests that this dispersal may have been followed by significant exchange of migrants.

In the case of *C. monoculus*, two different optimizations and presumed scenarios of dispersal and vicariance were reconstructed. The main difference between these scenarios is whether *C. monoculus* was ancestrally distributed in the Orinoco (i.e. already present there at the time it diverged from *C. ocellaris*), or more recently dispersed into the Orinoco. Evidence from the species' distribution, intraspecific mtDNA genealogy, and coalescent analysis clearly agrees with the latter inference (a young distribution in the Orinoco, Fig. 6a and c). First, *C. monoculus* has a relatively limited distribution in the Orinoco, as it is found only between Puerto Ayacucho and the Casiquiare. This limited distribution, despite wide habitat availability throughout the Orinoco (Winemiller 2001), is suggestive of a recent arrival in the basin. Second, if *C. monoculus* had been ancestrally distributed in both the Amazonas and Orinoco, barring demographic bottlenecks, we would expect to see much greater genetic diversity in the Orinoco basin. In contrast, only one of the 50 haplotypes observed in *C. monoculus* was found in the Orinoco basin, and this was shared with Amazonas fishes. As a result, it is not surprising that IMa estimated a very recent divergence between Orinoco and Amazonas (Negro) populations, using either the no migration or the migration model.

Finally, for *C. orinocensis*, DIVA was equivocal, providing three different possible scenarios: (1) dispersal from the Orinoco to Amazonas, (2) inheritance of an ancestral range that included both basins, or (3) dispersal from the Amazonas to Orinoco. Phylogeographic data provide some support for the former hypothesis. Haplotype diversity is higher in the Orinoco (although sampling effort may bias this pattern), and Orinoco haplotypes are relatively basal in the tree. Thus, we tentatively suggest that dispersal from the Orinoco to Amazonas is more likely than the reverse.

Considering the evidence from both intraspecific analyses and DIVA, we suggest that the best-supported biogeographical hypothesis for *Cichla* is depicted in Fig. 6a.

This scenario includes two dispersals of species from the Amazonas to the Orinoco, and a single dispersal from the Orinoco to the Amazonas. Importantly, in all cases, even the oldest IMa estimates of divergence time between Amazonas and Orinoco populations (<2 Ma) are considerably younger than estimates of the Amazonas-Orinoco vicariance event (~10 Ma, Lundberg *et al.* 1998b). This further supports the idea that the distributions of our focal species are the product of dispersal events rather than inherited ranges corresponding to the ancient paleo-Amazonas. However, it is possible that the phylogeny of *Cichla* does exhibit some signal of the Amazonas-Orinoco vicariance. In Fig. 6a, the ancestor to clade B is distributed in the Amazonas and Orinoco basins before undergoing allopatric speciation (vicariance). Fixing this node at 10 Ma implies a control region mutation rate of ~2.8%/Myr, much less than the published rates for cichlid fishes. Determining if the Amazonas-Orinoco vicariance event separated ancestral species in this genus will require a time-calibrated multi-locus phylogenetic analysis of *Cichla*. In any case, we suggest that our results support the conclusion that the combined use of intra- and interspecific data is a powerful approach for testing complex historical biogeographic hypotheses.

Although our current dataset, derived from three different species of *Cichla*, provides strikingly similar conclusions, the effect of any biogeographic feature depends on the ecological characteristics of the species impacted by it (e.g. Bermingham & Martin 1998; Thacker *et al.* 2007; Burridge *et al.* 2008). In a detailed study of the fish communities and environments of the Casiquiare, Winemiller *et al.* (2008) inferred that the Casiquiare River acts as a selective filter, where only some species are able to make the transition from stained, acidic water in the Negro River to unstained, neutral water in the upper Orinoco River. While *Cichla* species exhibit distinct habitat preferences (Jepsen *et al.* 1997), it appears that the species we studied are able to disperse across less suitable habitats and utilize the Casiquiare as a dispersal corridor. However, the three *Cichla* species with cross-drainage distributions do not show identical phylogeographic patterns or demographic estimates, implying that the timing and functional use of the Casiquiare River is mediated by the ecology of each species. Thus, to understand the utility of the Casiquiare region for the Amazonian ichthyofauna as a whole, it will be important to assess the importance of this dispersal corridor for taxa with varying ecophysiological requirements.

Acknowledgements

References

Bermerguy RL, Sena Costa JB (1991) Considerações sobre a evolução do sistema de drenagem da Amazônia e sua relação com o arcabouço tectônico-estrutual. *Museo Paraense Emilio Gouldi, Serio Ciencias da Terra*, **3**, 75–97.

Bermingham E, Martin AP (1998) Comparative mtDNA phylogeography of Neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology*, **7**, 499–517.

Burridge CP, Craw D, Waters JM (2006) River capture, range expansion, and cladogenesis: the genetic signature of freshwater vicariance. *Evolution*, **60**, 1038–1049.

Burridge CP, Craw D, Waters JM (2007) An empirical test of freshwater vicariance via river capture. *Molecular Ecology*, **16**, 1883–1895.

Burridge CP, Craw D, Jack DC, King TM, Waters JM (2008) Does fish ecology predict dispersal across a river drainage divide? *Evolution*, **62**, 1484–1499.

Chernoff B, Machado-Allison A, Saul WG (1991) Morphology, variation and biogeography of *Leporinus brunneus* (Pices: Characiformes: Anostomidae). *Ichthyological Exploration of Freshwaters*, **1**, 295–306.

Cooper JD, Vitalis R, Waser PM, Gopurenko D, Hellgren EC, Gabor TM, DeWoody JA (2010) Quantifying male-biased dispersal among social groups in the collared peccary (*Pecari tajacu*) using analyses based on mtDNA variation. *Heredity*, **104**, 79–87.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39**, 783–791.

Freeman B, Nico LG, Osentoski M, Jelks HL, Collins TM (2007) Molecular systematics of Serrasalmidae: deciphering the identities of piranha species and unraveling their evolutionary histories. *Zootaxa*, **1484**, 1–38.

Gery J (1969) The fresh-water fishes of South America. In: *Biogeography and Ecology in South America* (eds Fittkau EJ, Illies J, Klinge H, Schwabe GH, Sioli H), pp. 828–848. Dr. W. Junk Publishers, The Hague.

Hansson B, Hasselquist D, Tarka M, Zehtindjiev P, Bensch S (2008) Postglacial colonisation patterns and the role of

isolation and expansion in driving diversification in a passerine bird. *PLoS ONE*, **3**, e2794.

Harrison R, Hastings A (1996) Genetic and evolutionary consequences of metapopulation structure. *Trends in Ecology and Evolution*, **11**, 180–183.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the Natinal Academy of Sciences USA*, **104**, 2785–2790.

Hoorn C (1994) An environmental reconstruction of the Palaeo-Amazon River system (middle-late Miocene, Nw Amazonia). *Palaeogeography Palaeoclimatology Palaeoecology*, **112**, 187–238.

Hoorn C, Guerrero J, Sarmiento GA, Lorente MA (1995) Andean tectonics as a cause for changing drainage patterns in Miocene northern South-America. *Geology*, **23**, 237–240.

Hubert N, Renno JF (2006) Historical biogeography of South American freshwater fishes. *Journal of Biogeography*, **33**, 1414–1436.

Jepsen DB, Winemiller KO, Taphorn DC (1997) Temporal patterns of resource partitioning among *Cichla* species in a Venezuelan blackwater river. *Journal of Fish Biology*, **51**, 1085–1108.

Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol*, **4**, 18.

Kullander SO, Ferreira EJG (2006) A review of the South American cichlid genus *Cichla*, with descriptions of nine new species (Teleostei : Cichlidae). *Ichthyological Exploration of Freshwaters*, **17**, 289–398.

Lacy RC (1987) Loss of genetic diversity from managed populations: interaction effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*, **1**, 143–158.

Layman CA, Winemiller KO (2004) Size-based responses of prey to piscivore exclusion in a species-rich neotropical river. *Ecology*, **85**, 1311–1320.

Lopez-Rojas H, Machado-Allison A, Mago-Leccia F (1978) Review of ecological studies in tropical fish Communities by R.H. *Lowe-McConnell. Copeia*, **1988**, 503–505.

Lovejoy NR, de Araújo MLG (2000) Molecular systematics, biogeography and population structure of Neotropical freshwater needlefishes of the genus *Potamorrhaphis*. *Molecular Ecology*, **9**, 259–268.

Lundberg JG (1997) Fishes of the La Venta Fauna: additional taxa, biotic and paleoenvironmental implications. In: *Vertebrate Paleontology in the Neotropics: The Miocene Fauna of La Venta Colombia* (eds Kay RF, Madden RH, Cifelli RL, Flynn JJ), pp. 67–91. Smithsonian Institution Press, Washington, DC.

Lundberg JG (1998a) The temporal context for the diversification of neotropical fishes. In:*Phylogeny and Classification of Neotropical Fishe* (eds Malabarba LR, Reis RE, Vari RP, Lucena CAS, Lucena ZMS). pp. 49–68, EDIPUCRS, Rio de Janeiro, Brazil.

Lundberg JG, Marshall LG, Guerrero J *et al.* (1998b) The stage for neotropical fish diversification: a history of tropical South American rivers. In: *Phylogeny and Classification of Neotropical Fishes* (eds Malabarba LR, Reis RE, Vari RP, Lucena CAS, Lucena ZMS), pp. 13–48. EDIPUCRS, Rio de Janeiro, Brazil.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.

Maddison W, Maddison DR (2008) Mesquite: a modular system for evolutionary analysis. Version 2.5. http://mesquiteproject.org.

Mago-Leccia F (1971) La ictiofauna del Casiquiare. *Revista Defense de la Naturaleza, Caracas*, **1**, 5–10.

Mäkinen HS, Merilä J (2008) Mitochondrial DNA phylogeography of the three-spined stickleback (*Gasterosteus aculeatus*) in Europe—Evidence for multiple glacial refugia. *Molecular Phylogenetics and Evolution*, **46**, 167–182.

Meyer A (1993) Evolution of mitochondrial DNA in fishes. In: *Biochemistry and Molecular Biology of Fishes Vol 2* (eds Hochachka PW, Mommsen P), pp. 1–38. Elsevier Press, Amsterdam, Netherlands.

Nylander JA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, **53**, 47–67.

Peters JL, Gretes W, Omland K (2005) Late Pleistocene divergence between eastern and western populations of wood ducks (*Aix sponsa*) inferred by the 'isolation with migration' coalescent method. *Molecular Ecology*, **14**, 3407–3418.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Reis RE, Kullander SO, Ferraris CJ (2003) *Checklist of the Freshwater Fishes of South and Central America*, p. 729. EDIPUCRS, Porto Alegre, Brazil.

Rice AH (1921) The Rio Negro, the Casiquiare Canal, and the Upper Orinoco, September 1919-April 1920. *The Geographic Journal*, **58**, 321–343.

Ronquist F (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, **46**, 195–203.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Sioli H (1984) The Amazon and its main affluents: hydrography, morphology of the river courses, and river types. In: *The Amazon: Limnology and Landscape Ecology of a Mighty Tropical River and its Basin* (ed. Sioli H), pp. 127–163, Dr. W. Junk Publishers, The Hague.

Stern K (1970) Der Casiquiare-Kanal, einst und jetzt. *Amazoniana*, **2**, 401–416.

Strasburg JR, Rieseberg RH (2009) How robust are ''Isolation with Migration'' analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution*, **27**, 297–310.

Sturmbauer C, Baric S, Salzburger W, Ruber L, Verheyen E (2001) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular Biology and Evolution*, **18**, 144–154.

Swofford DL (2000) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thacker CE, Unmack PJ, Matsui L, Rifenbark N (2007) Comparative phylogeography of five sympatric *Hypseleotris* species (Teleostei: Eleotridae) in south-eastern Australia reveals a complex pattern of drainage basin exchanges with little congruence across species. *Journal of Biogeography*, **34**, 1518–1533.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Turner TF, McPhee MV, Campbell P, Winemiller KO (2004) Phylogeography and intraspecific genetic variation of prochilodontid fishes endemic to rivers of nothern South America. *Journal of Fish Biology*, **64**, 186–201.

Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science*, **300**, 325–329.

Wahlund S (1928) Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, **11**, 65–106.

Weitzman SH, Weitzman M (1982) Biogeography and evolutionary diversification in neotropical freshwater fishes, with comments on the refuge theory. In: *Biological Diversification in the Tropics* (ed. Prance GT), pp. 403–422. Columbia University Press, New York, NY.

Willis SC, Nunes MS, Montana CG, Farias IP, Lovejoy NR (2007) Systematics, biogeography, and evolution of the neotropical peacock basses *Cichla* (Perciformes : Cichlidae). *Molecular Phylogenetics and Evolution*, **44**, 291–307.

Winemiller KO (2001) Ecology of peacock cichlids (*Cichla* spp.) in Venezuela. *Journal of Aquariculture & Aquatic Sciences*, **9**, 93–112.

Winemiller KO, Jepsen DB (1998) Effects of seasonality and fish movement on tropical river food webs. *Journal of Fish Biology* **53** (Supplement A), 267–296.

Winemiller K, Willis SC (in press) Biogeography of the Vaupes Arch and Casiquiare River: barriers and passages between the Amazon and Orinoco. Ch. 14 In: *Historical Biogeography of Neotropical Freshwater Fishes* (eds Albert J, Reis R). University of California Press, Berkeley, CA.

Winemiller KO, López-Fernández H, Taphorn DC, Nico L, Barbarino-Duque A (2008) Fish assemblages of the Casiquiare River, a corridor and zoogeographical filter for dispersal between the Orinoco and Amazon basins. *Journal of Biogeography*, **35**, 1551–1563.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Sampled localities and approximate distributions of the described species of *Cichla* (*C. nigromaculata* is considered synonymous with *C. monoculus*). Locality abbreviations follow Table S2.

Table S1 Sampling localities and sample sizes for the focal species (control region).

Table S2 Sampling localities and sample sizes for the phylogenetic analysis (cytochrome *b*, ATPase 8,6, and control region).

Table S3 Statistics from the marginal distributions of parameters from the final IMa analyses of *C. temensis*, *C. orinocensis*, and *C. monoculus*.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.